Sourabh Dilip Powar.

Q1 a) Likelihood is how likely the output y from $P(y/x)$ captured by hypothesis $h(x)$

$$p(y/x) = \begin{cases} h(x) & \text{if } y = 1 \\ 1-h(x) & \text{if } y = -1 \end{cases} \quad -①$$

Since the data points are independently generated the probability of getting all thy y_n's in dataset from x_n's will be.

$$E_{in}(w) = \prod_{n=1}^{N} P(y_n/x_n) \quad -②$$

To minimize above equation take log & multiply by (-1)

$$\min(E_{in}(w)) = (-1)\log \prod_{n=1}^{N} P(y_n/x_n)$$

$$= (-1) \log[(P(y_1/x_1)(P(y_2/x_2)\dots P(y_n/x_n)]$$

$$= (-1)[\log(P(y_1/x_1)) + \log(P(y_2/x_2)) \dots + \log(P(y_n/x_n))$$

$$= \log(P(y_1/x_1))^{-1} + \log P(y_2/x_2)^{-1} \dots \log P(y_n/x_n)^{-1}$$

$$= \sum_{n=1}^{N} \ln \frac{1}{p(y_n/x_n)} \quad -③$$

For given hypothesis function.

$$E_{in}(w) = \sum_{n=1}^{N} [\![y_n = +1]\!] \ln \frac{1}{p(1/x)} + [\![y_n = -1]\!] \ln \frac{1}{p(-1/x)}$$

$$= \sum_{n=1}^{N} [\![y_n = +1]\!] \frac{1}{h(x_n)} + [\![y_n = -1]\!] \ln \frac{1}{1-h(x_n}$$

Q1b) For case $h(x) = \theta(w^T x)$ where $\theta(s) = \frac{1}{1+e^{-s}}$

From equation ② in above question

$$\min(E_{in}(w)) = \sum_{n=1}^{N} \frac{1}{p(y_n/x_n)} = \cancel{\sum_{n=1} \frac{1}{\theta(s)}}$$

To minimize divide by $1/N$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{1}{p(y_n/x_n)}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\theta(y_n w^T x_n)}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\frac{1}{1+e^{-(y_n w^T x_n)}}}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left(1+e^{-y_n w^T x_n}\right)$$

Q2a) $E_2(w) = \|Xw - y\|^2 + \lambda \|w\|^2$

$\left[\because \|x\| = (x^T x)^{1/2}\right] = (Xw-y)^T (Xw-y) + \lambda w^T w^2$

$$= \left((Xw)^T - y^T\right)(Xw-y) + \lambda w^T w^2$$

$$= \left((w^T X^T) - y^T\right)(Xw-y) + \lambda w^T w^2$$

$$= w^T X^T X w - w^T X^T y - y^T X w + y^T y + \lambda w^T w$$

Now $y^T X w = w^T X^T y \Rightarrow$ cyclic Transpose

$$E_{in} = w^T X^T X w - 2 w^T X^T y + y^T y + \lambda w^T w^2$$

Rules:

$$\nabla (w^T A w) = (A + A^T) w$$
$$\nabla (w^T b) = b$$

$$\nabla E_{in}(w) = (X^T X + (X^T X)^T) w - 2 X^T y + 0 + 2\lambda w$$
$$= 2 (X^T X - X^T y + \lambda I) w$$

For minimum w,

$$(X^T X w - X^T y + \lambda I w) = 0$$
$$\therefore \quad X^T X w + \lambda I w = X^T y$$
$$\therefore \quad (X^T X + \lambda I) w = X^T y$$

multiply by inverse on both side

$$\underline{\underline{w = (X^T X + \lambda I)^{-1} X^T y}}$$

Q 2 b) Now conside Matrix

$$X^T X + \lambda I$$

1) That $X^T X$ is positive semidefinite.

proof.

$u^T X^T X u = \|Xu\|^2 \geq 0$ so all eigen valuse must be ≥ 0

2) The eigenvalues of $X^T X + \lambda I$ are $\mu_i + \lambda$ where μ_i are eigenvalues of $X^T X$.

All eigenvalues are strictly positive So it must be invertible.

Scanned by CamScanner

Q3 a)

$$E(\omega) = \frac{1}{N}\sum_{n=1}^{N} \ln(1+e^{-y_n \omega^T x_n})$$

$$\therefore \nabla E(\omega) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{1+e^{-y_n w^T x_n}} \frac{\partial}{\partial w}(1+e^{-y_n w^T x_n})$$

— chain rule. $\frac{\partial}{\partial x}\log x = \frac{1}{x}$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{1}{1+e^{-y_n w^T x_n}} \times e^{-y_n w^T x_n} \times \frac{\partial}{\partial w}(-y_n w^T x_n)$$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{-y_n x_n}{(1+e^{(-y_n w^T x_n)}) \times (e^{y_n w^T x_n})}$$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{-y_n x_n}{e^{y_n w^T x_n}+1}$$

$$= -\frac{1}{N}\sum_{n=1}^{N} \frac{y_n x_n}{1+e^{y_n w^T x_n}}$$

Q3 b) For given $\theta(x) = \frac{1}{1+e^{-x}}$

predicted class of $x = \begin{cases} 1 & \theta(\omega^T x) \geq 0.5 \\ -1 & \theta(\omega^T x) < 0.5 \end{cases}$

Suppose $p(y=1/x)=1$ as $\theta(\omega^T x) \geq 0.5$.

$$\therefore \frac{1}{1+e^{-\omega^T x}} \geq 0.5 \quad \text{— (1)}$$

$$\therefore 2 \geq 1+e^{-\omega^T x}$$

$$\therefore \quad e^{-w^T x} = 1$$

Taking log.

$$-w^T x = 0$$ → This in linear equation in x

$$\therefore \sum_{n=0}^{1} x_n w_n = 0 \qquad x_0 = 1$$

So the decision boundaries of logistic regression are linear though $\theta(x)$ is not linear.

Q3c) For change in threshold of class x.

predicted. class of x $\begin{cases} 1 & \text{if } \theta(w^T x) \geq 0.9 \\ -1 & \text{if } \theta(w^T x) \leq 0.9. \end{cases}$

From ① $p(y=1/x) = 1$ if

$$\theta(w^T x) \geq 0.9.$$

$$\frac{1}{1+e^{w^T x}} \geq 0.9$$

$$\therefore \frac{1}{0.9} \geq 1 + e^{w^T x}$$

$$\therefore \frac{1}{0.9} - 1 \geq e^{w^T x}$$

$$\therefore \frac{0.1}{0.9} \geq e^{w^T x}$$

$$\therefore \log\left(\frac{1}{9}\right) \geq w^T x$$ — This is also a linear equation

$\omega^T x - k_0 = 0$ — linea equation with
$\omega_0 = k$.

Q3d) The important property is the bound of the classification based on the threshold.
This bound will have the decision boundaries which are the range based on the probability.

$$p(y/x) = \begin{cases} 1 & \theta(x) \geq \text{threshold} \\ -1 & \theta(x) < \text{threshold} \end{cases}$$

Thus the values are bounded by $y=1$ & $y=-1$
If we change bounds to linear or exponential then the boundaries will change to non-linear.

Decision boundary is property of hypothesis.